Real World Real-time Automatic Recognition of Facial Expressions

Ying-li Tian Lisa Brown Arun Hampapur Sharat Pankanti Andrew Senior and Ruud Bolle
Exploratory Computer Vision Group, IBM T. J. Watson Research Center
PO Box 704, Yorktown Heights, NY 10598
{yltian,lisabr,arunh,sharat,aws,bolle}@us.ibm.com

Abstract

Most facial expression analysis systems attempt to recognize facial expressions from data collected in a highly controlled laboratory with very high resolution frontal faces (face regions greater than 200 x 200 pixels) and cannot handle large head motions. In real environments such as smart meetings, a facial expression analysis system must be able to automatically recognize expressions at lower resolution and handle the full range of head motion. This paper describes a real-time system to automatically recognize facial expressions in relatively low resolution face images (around 50x70 to 75x100 pixels). To handle the full range of head motion, we detect the head instead of the face. Then the head pose is estimated based on the detected head. For frontal and near frontal views of the face, the location and shape features are computed for expression recognition. Our system successfully deals with complex real world interactions, as demonstrated on the PETS2003 dataset.

1 Introduction

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. Automatic facial expression analysis has attracted the interest of many computer vision researchers. In the past decade, much progress has been made to build computer systems to understand and use this natural form of human communication from either static facial images or image sequences [2, 4, 5, 7, 12, 13, 15, 20, 22, 24].

Current approaches to automated facial expression analysis typically attempt to recognize a small set of prototypic emotional expressions, i.e. joy, surprise, anger, sadness, fear, and disgust. Suwa *et al.* [19] presented an early attempt to analyze facial expressions by tracking the motion of twenty identified spots on an image sequence. Essa and Pentland [7] developed a dynamic parametric model based on a 3D geometric mesh face model to recognize 5 prototypic expressions. Mase [13] manually selected facial regions that corresponded to facial muscles and computed motion within these regions using optical flow. The work by Yacoob and Davis [22] used optical flow like Mase's work, but tracked the motion of the surface regions of facial features (brows, eyes, nose, and mouth) instead of the motion of the underlying muscle groups. Zhang [24] investigated the use of two types of facial features: the geometric positions of 34 fiducial points on a face and a set of multi-scale, multi-orientation Gabor wavelet coefficients at these points for facial expression recognition.

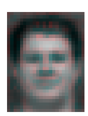
Several systems [4, 5, 8, 12, 21] attempt to recognize fine-grained changes in facial expression based on the Facial Action Coding System (FACS) which developed by Ekman and Friesen [6] for describing facial expressions by action units (AUs). Of 44 FACS AUs that they defined, 30 AUs are anatomically related to the contractions of specific facial muscles: 12 are for upper face, and 18 are for lower face. AUs can occur either singly or in combination. Automatic recognition of FACS action units (AU) is a difficult problem, and relatively little work has been reported. AUs have no quantitative definitions and, as noted, can appear in complex combinations.

Mase [13] and Essa [7] described patterns of optical flow that corresponded to several AUs, but did not attempt to recognize them. Bartlett et al. [1] and Donato et al. [5] reported experimental results of upper and lower face AU recognition. They both used image sequences that were free of head motion, manually aligned faces using three coordinates, rotated the images so that the eyes were in horizontal, scaled the images, and, finally, cropped a window of 60x90 pixels for upper face and lower face respectively. They recognized 6 single upper face AUs and 2 lower face AUs and 4 AU combinations. For analysis purpose, they treated each combination as if it were a separate new AU. Cohn *et al.* [4] and Lien *et al.* [12] used dense-flow, feature-point tracking, and edge extraction to recognize 4 upper face AUs and 2 combinations and 4 lower face AUs and 5 combinations. Again each AU combination was regarded as a separate new AU. Tian *et al.* [21] reported some of the most extensive experimental results on AU recognition. Their system can automatically detect the face [16] and cope with a large change of appearance and limited out-of-plane head motion. No image alignment was necessary, and in-plane and limited out-of-plane head motion can be handled. To increase the robustness and accuracy of the feature extraction, multi-state face-component models were devised. The system recognizes 16 of the 30 AUs whether they occur alone or in combinations.

The limitations of the existing systems are summarized as following:

- Most systems attempt to recognize facial expressions from data collected in a highly controlled laboratory with very high resolution frontal faces (face regions greater than 200 x 200 pixels).
- Most system need some manual preprocessing.
- Most systems cannot handle large out-of-plane head motion.
- None of these systems deals with complex real world interactions.
- Except the system proposed by Moses *et al.* [14], none of those systems performs in real-time.

Table 1. A face at different resolutions. All images are enlarged to the same size. At 48x64 pixels the facial features such as the corners of the eyes and the mouth become hard to detect. Facial expressions are not recognized at 24x32 pixels.

Face Process	 96x128	 **69x93**	 48x64	 24x32
Detect?	Yes	Yes	Yes	Yes
Pose?	Yes	Yes	Yes	Yes
Recognize?	Yes	Yes	Yes	Maybe
Features?	Yes	Yes	Maybe	No
Expressions?	Yes	Yes	Maybe	No

In this paper, we propose an expression recognition system which addresses many of these limitations. In real environments, a facial expression analysis system must be able to:

- fully automatically recognize expressions.
- handle a full range of head motions.
- recognize expressions in face images with relatively lower resolution.
- recognize expressions in lower intensity.
- perform in real-time.

Table 1 shows a face at different resolutions. Most automated face processing tasks should be possible for a 69x93 pixel image. At 48x64 pixels the facial features such as the corners of the eyes and the mouth become hard to detect. The facial expressions may be recognized at 48x64 and are not recognized at 24x32 pixels. This paper describes a real-time system to automatically recognize facial expressions in relatively low resolution face images (50x70 to 75x100 pixels). To handle the full range of head motion, instead of detecting the face, the head pose is estimated based on the detected head. For frontal and near frontal views of the face, the location and shape features are computed for expression recognition. Our system successfully deals with complex real world interactions, as demonstrated on the PETS2003 dataset. Section 2 presents the overall architecture of the system. Its components: background subtraction, head detection and head pose estimation are presented in Sections 3, 4 and 5 respectively. Section 6 describes the method for facial feature extraction and tracking. Section 7 discusses our method for recognizing expressions. Section 8 presents results on the PETS test sequences. We summarize our paper and present future directions in Section 9.

2 System architecture

In this paper we describe a new facial expression analysis system designed to automatically recognize facial expressions in real-time and real environments, using relatively low resolution face images. Figure 1 shows the structure of the tracking system.

The input video sequence is used to estimate a background model, which is then used to perform background subtraction, as described in Section 3. In Section 4, the resulting foreground regions are used to detect the head. After finding the head, head pose estimation is performed to find the head in frontal or near-frontal views. The facial features are extracted only for those faces in which both eyes and mouth corners are visible. The normalized facial features are input to a neural network based expression classifier to recognize different expressions.

In this paper we describe results using the PETS 2003 evaluation dataset scenario A, camera 1. In this scenario, each person enters the conference room one after the other, goes to his places, presents himself to the frontal camera, and sit down. Then each person looks at the other people with different expressions. For reasons of speed and storage economy, we have chosen to process the video at half resolution for head detection and head pose estimation. The system operates on AVI video files (Cinepak compressed) generated from the distributed JPEG images. Figure 2(a), (b), and (c) show the face resolution of 3 subjects

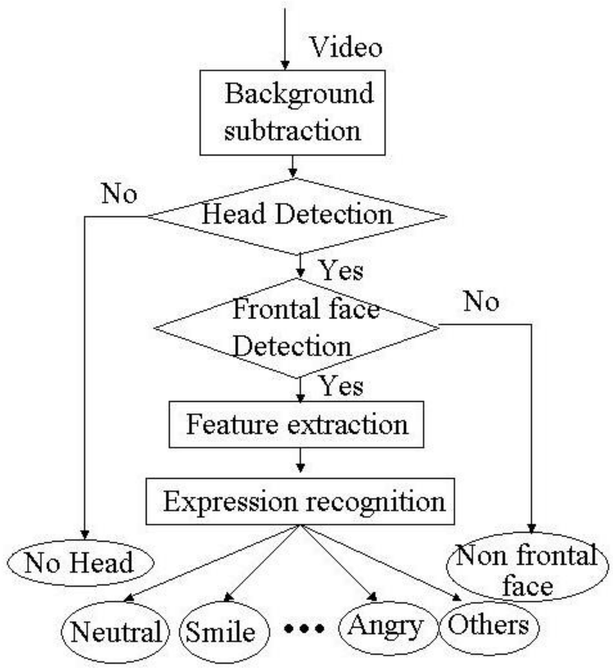


Figure 1. Block diagram of the expression recognition system

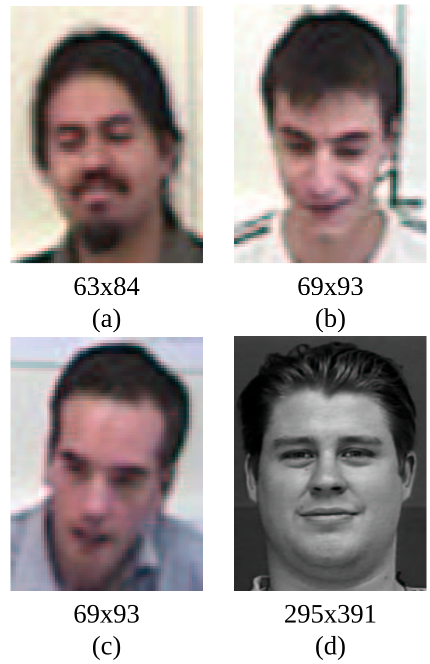


Figure 2. Face resolution from the original JPEG image in PETS4 Dataset. For display purpose, the face region is cropped ((a), (b), (c)). Notice the resolution difference; the images in PETS4 dataset are very blurry because of the compression. For comparison, a face image from Cohn-Kanade AU-Coded Face Expression Image Database is shown in (d).

in the dataset. The face region is cropped from the original JPEG image. Notice the resolution difference; the images in PETS4 dataset are very blurry because of the compression. For comparison, the face region from Cohn-Kanade AU-Coded Face Expression Image Database is shown in 2(d).

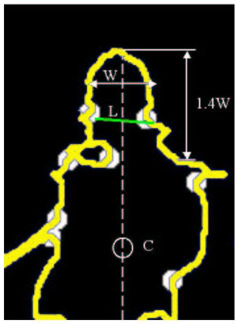
3 Background estimation and subtraction

The background subtraction approach presented here is similar to that taken by Horprasert *et al.* [10] and is an attempt to make the background subtraction robust to illumination changes. The background is modelled statistically at each pixel. The estimation process computes the brightness distortion and color distortion in RGB color space. Each pixel i is modelled by a 4-tuple (E_i, s_i, a_i, b_i), where E_i is a vector with the means of the pixel's red, green, and blue components computed over N background frames; s_i is a vector with the standard deviations of the color values; a_i is the variation of the brightness distortion; and b_i is the variation of the chromaticity distortion. We have also developed an active background estimation method that can deal with moving objects in the frame. First, we calculate image difference over three frames to detect the moving objects. Then the statistical background model is constructed, excluding these moving object regions.

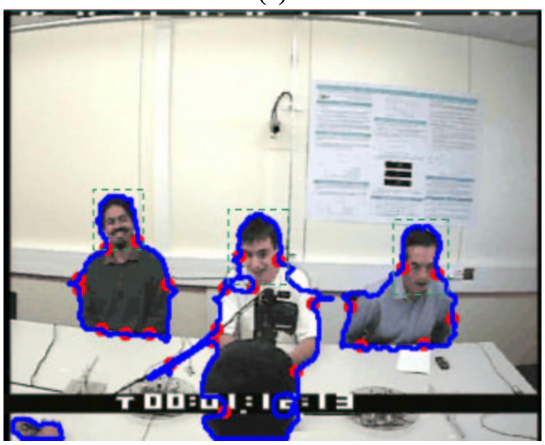
By comparing the difference between the background image and the current image, a given pixel is classified into one of four categories: original background, shaded background or shadow, highlighted background, and foreground objects. The details of automatic threshold selection can be found in the original paper [10]. Finally, a morphology step is applied to remove small isolated spots and fill holes in the foreground image.

4 Head detection

In order to handle the full range of head motion, we detect the head instead of detecting the face. The head detection uses the smoothed silhouette of the foreground object as segmented using background subtraction. Based on human intuition about the parts of an object, a segmentation into parts generally occurs at the *negative curvature minima* (NCM) points of the silhouette [9] as shown with small circles in Figure 3. The boundaries between parts are called *cuts* (shown as the line *L* in Figure 3(a)). Singh *et al.* [18] noted that human vision prefers the partitioning scheme which uses the shortest cuts. They proposed a short-cut rule which requires a cut: 1) be a straight line, 2) cross an axis of local symmetry, 3) join two points on the outline of a silhouette and at least one of the two points is NCM, 4)



(a)



(b)

Figure 3. Head detection steps. (a) calculate the cut of the head part. (b) obtain the correct head region from the cut of the head part.

be the shortest one if there are several possible competing cuts.

In our system, the following steps are used to calculate the cut of the head:

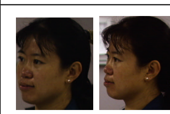
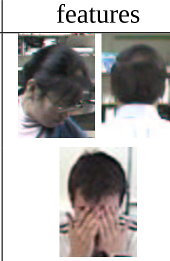
- calculate the contour centroid C and the vertically symmetry axis y of the silhouette.
- compute the cuts for the NCMs which are located above the contour centroid C.
- measure the salience of a part's protrusion which is defined as the ratio of the perimeter of the part (excluding the cut) to the length of the cut.
- test if the salience of a part exceeds a low threshold.
- test if the cut crosses the vertical symmetry axis y of the silhouette.
- select the top one as the cut of the head if there are several possible competing cuts.

After the cut of the head L is detected, the head region can be easily determined as the part above the cut. As shown in Figure 3(b), in most situations, only part of the head lies above the cut. To obtain the correct head region, we first calculate the head width W, then the head height H is enlarged to $\alpha * W$ from the top of the head. In our system, $\alpha = 1.4$.

5 Head pose detection

After the head is located, the head image is converted to gray-scale, histogram equalized and resized to the estimated resolution. Then we employ a three layer neural network (NN) to estimate the head pose. The inputs to the network are the processed head image. The outputs are the head poses. Unlike our previous work [3] which trained the NN to 9 pan angles from -90 to +90 degrees in steps of 22.5 degrees, here only 3 head pose classes are trained for expression analysis: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded face. Table 2 shows the definitions and some examples of the 3 head pose classes. In the frontal or near frontal view, both eyes and lip corners are visible. In side view or profile, at least one eye or one corner of the mouth becomes self-occluded because of the head turn. All other reasons cause more facial features to not be detected such as the back of the head, occluded face, and face with extreme tilt angles are treated as one class.

Table 2. The definitions and examples of the 3 head pose classes: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded faces.

Poses	Frontal near frontal	Side view profile	Others
Definitions	Both eyes and lip corners are visible	One eye or one lip corner is occluded	Not enough facial features
Examples			

To train the NN, we have used the CMU Pose, Illumination, and Expression (PIE) Database of Human Faces for our ground truth data [17]. This database contains images of 68 people under 13 poses, 43 different illumination conditions and 4 different expressions. In our study, we only use 9 poses of neutral expressions, from -90 to +90 degrees about the vertical axis and natural room lighting. In addition to the CMU PIE database, we collected the images of the back of the head, occluded face, and face with extreme tilt angles in our lab as ground truth data for the 3rd class.

6 Facial feature extraction for frontal or near-frontal face

After estimating the head pose, the facial features are extracted only for the face in the frontal or near-frontal view. Since the face images are in relatively low resolution in most real environments, the detailed facial features such as the corners of the eyes and the upper or lower eyelids are not available (see the face examples in Figure 2(a), (b), and (c)). To recognize facial expressions, however, we need to detect reliable facial features. We observe that most facial feature changes that are caused by an expression are in the areas of eyes, brows and mouth. In this paper, two types of facial features in these areas are extracted: location features and shape features.

6.1 Location feature extraction

In our system, six location features are extracted for expression analysis. They are eye centers (2), eyebrow inner endpoints (2), and corners of the mouth (2). The feature extraction approach presented here is similar to that taken by Yang *et al.* [23] and is an attempt to make the extraction of the eye centers more accurate and robust.

(a) (b) (c) (d)

Figure 4. Iterative thresholding of the face to find eyes and brows. (a) grey-scale face image, (b) threshold = 30, (c) threshold = 42, (d) threshold = 54.

Eye centers and eyebrow inner endpoints: To find the eye centers and eyebrow inner endpoints inside the detected frontal or near frontal face, we have developed an algorithm that searches for two pairs of dark regions which correspond to the eyes and the brows by using certain geometric constraints such as position inside the face, size and symmetry to the facial symmetry axis. Similar to paper [23], the algorithm employs an iterative thresholding method to find these dark regions under different or changing lighting conditions.

Figure 4 shows the iterative thresholding method to find eyes and brows. Generally, after *five* iterations, all the eyes and brows are found. If satisfactory results are not found after 20 iterations, we think the eyes or the brows are occluded or the face is not in a near frontal view. Unlike the work of Yang *et al.* to find one pair of dark regions for the eyes only, we find two pairs of parallel dark regions for both the eyes and eyebrows. By doing this, not only are more features obtained, but also the accuracy of the extracted features is improved. As shown in Figure 4(b), the right brow and the left eye is wrongly extracted as the two eyes in Yang's approach. Figure 4(d) shows the correct positions are extracted for all the eyes and eyebrows in our method. Then the eye centers and eyebrow inner endpoints can be easily determined. If the face image is continually in the frontal or near frontal view in an image sequence, the eyes and brows can be tracked by simply searching for the dark pixels around their positions in the last frame.

Mouth corners: After finding the positions of the eyes, the location of the mouth is first predicted. Then the vertical position of the line between the lips is found using an integral projection of the mouth region proposed by Yang *et al.* [23]. Finally, the horizontal borders of the line between the lips is found using an integral projection over an edge-image of the mouth. After Yang *et al.* , the following steps are use to track the corners of the mouth: 1) Find two points on the line between the lips near the previous positions of the corners in the image 2) Search along the darkest path to the left and right, until the corners are found. Finding the points on the line between the lips can be done by searching for the darkest pixels in search windows near the previous mouth corner positions. Because there is a strong change from dark to bright at the location of the corners, the corners can be found by looking for the maximum contrast along the search path. The details of the tracking method of the mouth corners can be found in the original paper [23].

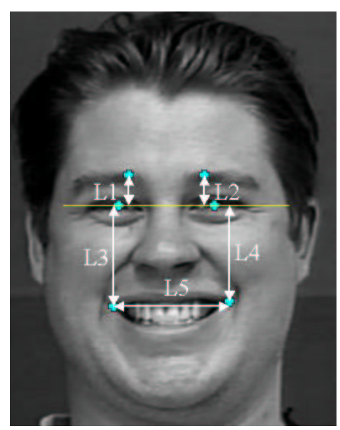


Figure 5. Face location feature representation for expression recognition.

6.2 Location feature representation

After extracting the location features, the face can be normalized to a canonical face size based on two of these features, i.e., the eye-separation after the line connecting two eyes (*eye-line*) is rotated to horizontal. In our system, all faces are normalized to 90 $\times$ 90 pixels by re-sampling. We transform the extracted features into a set of parameters for expression recognition. We represent the face location features by 5 parameters, which are shown in Figure 5. These parameters are the distances between the *eye-line* and the corners of the mouth, the distances between the *eye-line* and the inner eyebrows, and the width of the mouth (the distance between two corners of the mouth).

6.3 Shape feature extraction

Another type of distinguishing feature is the shape of the mouth. Global shape features are not adequate to describe the shape of the mouth. Therefore, in order to extract the mouth shape features, an edge detector is applied to the normalized face to get an edge map. This edge map is divided into 3 $\times$ 3 zones as shown in Figure 6(b). The size of the zones is selected to be half of the the distance between the eyes. The mouth shape features are computed from zonal shape histograms of the edges in the mouth region.

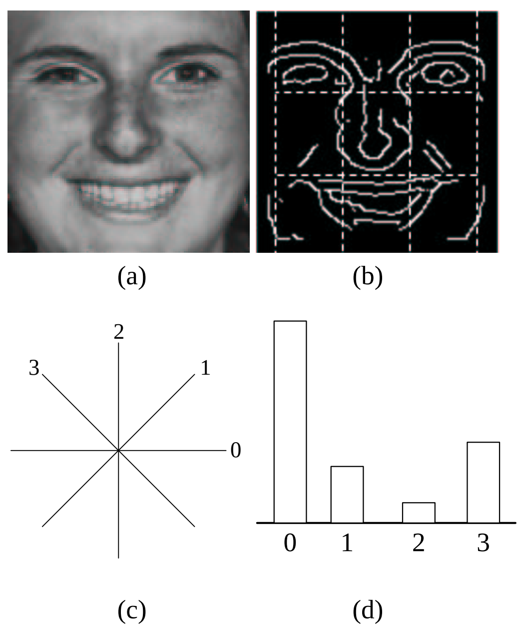


Figure 6. Zonal-histogram features. (a) normalized face, (b) zones of the edge map of the normalized face, (c) four quantization levels for calculating histogram features, (d) histogram corresponding to the middle zone of the mouth.

To place the 3 $\times$ 3 zones onto the face image, the centers of the top-left and top-right zone are placed at the locations of the eyes . In the vertical direction, then, the face is divided into three equal parts which correspond to the eyes and eyebrows region, the nose region and the mouth region, respectively.

The coarsely quantized edge directions are represented as local shape features and more global shape features are presented as histograms of local shape (edge directions) along the shape contour. The edge directions are quantized into 4 angular segments (Figure 6(c)). Representing the whole mouth as one histogram does not capture the local shape properties that are needed to distinguish facial expressions. Therefore we use the zones to compute three histograms of the edge directions. In the current system, only the three zones in the mouth region are used to obtain shape histograms. Hence, the mouth is represented as a feature vector of 12 components (3 histograms of 4 components). An example of the histogram of edge directions corresponding to the middle zone of the mouth is shown in Figure 6(d).

7 Expression recognition

We used a neural network-based recognizer having the structure shown in Figure 7. The standard back-propagation in the form of a three-layer neural network with one hidden layer was used to recognize facial expressions. The inputs to the network were the 5 location features (Figure 5) and the 12 zone components of shape features of the mouth (Figure 6). Hence, a total of 17 features were used to represent the amount of expression in a face image. The outputs were a set of expressions. In our system, 5 expressions were recognized. They were neutral, smile, angry, surprise, and others (including fear, sad, and disgust). We tested various numbers of hidden units and found that 6 hidden units gave the best performance.

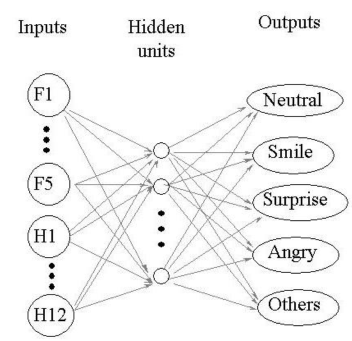


Figure 7. Neural network-based recognizer for facial expressions.

A total of 1088 images from the first portion of the Cohn-Kanade AU-Coded Face Expression Image Database [11] were used to train the network. The subjects in the first portion of the Cohn-Kanade database are 100 university students enrolled in introductory psychology classes. The demographics of this population are 65% female and 35% male, and 82% Caucasian, 15% African and 3% Asian or Latino ranging ranging in age from 18 to 30 years. The subjects were instructed by an experimenter to perform expressions. Subjects' facial behavior was recorded in an observation room using a frontal camera. The image sequences begin with a neutral face and are digitized into 640 pixel arrays with 8-bit gray-scale.

8 Experimental results

The goal of our effort was to develop a real-time expression recognition system for handling the full range of head motion in low resolution images. Given this focus, we report results on PETS dataset Scenario A, camera 1. Without the absolute neutral face for each subject, it is very difficult to recognize the expressions at very low intensity. It is difficult even for human beings. We showed 4 images with the expressions in low intensity to 5 of our group members; no one matched results to the ground truth which was provided by the PETS dataset (original GT). Based on the majority, we generated a new ground truth (modified GT) with head pose from frame 118000 to frame 118190.

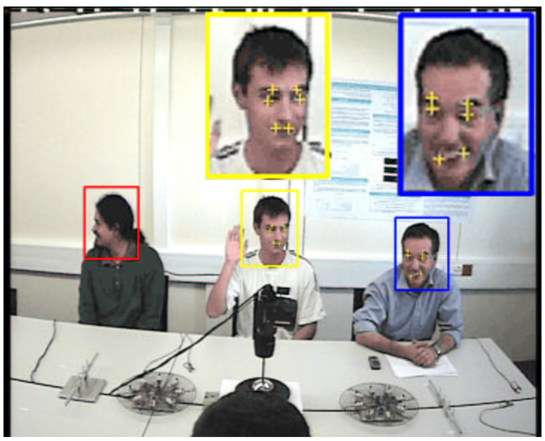


Figure 8. Head position detection and location feature extraction result on frame 118172.

Figure 8 shows the head position detection and location feature extraction result on frame 118172. Table 3 shows the recognition results for these frames by comparing to the *modified GT*. The head detection and head pose estimation worked well. For the expression recognition, many frames with very low intensity expressions were recognized as neutral. Since in our training data, most surprise expression images are with large opened mouth and large eyebrow motion, 4 frames with large eyebrow raised up in the test data are wrongly recognized as surprise. Some frames are recognized as other expressions because of the combination of talking when subject 4 is smiling. Table 4 shows some examples of recognized expressions.

9 Summary and conclusions

Automatically recognizing facial expressions is important to understand human emotion and paralinguistic communication, to design multimodal user interfaces, and to related applications such as human identification. It is very challenging to develop a system that can perform in real time and in real world because of low image resolution, low expression intensity, and the full range of head motion. We have developed an automatic expression recognition system that addresses all the above challenges and successfully deals with complex real world interactions, as demonstrated on the PETS2003 dataset. The results show that without the absolute neutral face for each subject, it is very difficult to recognize the expressions in very low intensity. In most real word interactions, the facial feature changes are caused by both talking and expression changes. More accurate results will be achieved if the system can detect talking in a higher level and only performs the expression recognition procedure for those images only content expression changes. We feel that further efforts will be required for distinguishing talking and expression changes by fusing audio signal processing and visual image analysis. Also it will benefit the expression recognition accuracy by using the sequential information instead of using each frame.

Table 3. Results of expression recognition and head pose estimation based on the *modified GT* which was generated by the authors.

		Test result				
		Non frontal	Neutral	Smile	Surprise	Others
GT	Non frontal	193	0	1	0	0
	Neutral	0	32	0	0	0
	Smile	0	22	313	4	5

Table 4. Examples of recognized expressions. *Original GT* means that the ground truth provided by PETS4 dataset. *Modified GT* means that the ground truth was generated by the authors.

					
Original GT	smile	smile	smile	smile	smile
Modified GT	profile	sideview	smile	smile	neutral
Test result	profile	sideview	surprise	others	neutral

References

[1] M.S. Bartlett, J.C. Hager, P.Ekman, and T.J. Sejnowski. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36:253–264, 1999.

[2] M. J. Black and Y. Yacoob. Trcking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Proc. Of International conference on Computer Vision*, pages 374–381, 1995.

[3] Lisa Brown and Ying li Tian. Comparative study of coarse head pose estimation. In *IEEE Workshop on Motion and Video Computing, Orlando FL, Dec. 5-6*, 2002.

[4] J. F. Cohn, A. J. Zlochower, J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. *Psychophysiology*, 36:35–43, 1999.

[5] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.

[6] P. Ekman and W.V. Friesen. *The Facial Action Coding System: A Technique For The Measurement of Facial Movement*. Consulting Psychologists Press, Inc., San Francisco, CA, 1978.

[7] I.A. Essa and A.P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 19(7):757–763, July 1997.

[8] B. Fasel and J. Luttin. Recognition of asymmetric facial action unit activities and intensities. In *Proceedings of International Conference of Pattern Recognition*, 2000.

[9] D.D. Hoffman and W.A. Richards. Parts of recognition. *Cognition*, 18:65–96, 1984.

[10] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *IEEE ICCV'99 Frame-Rate Workshop*, 1999.

[11] T. Kanade, J.F. Cohn, and Y.L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of International Conference on Face and Gesture Recognition*, pages 46–53, March, 2000.

[12] Jenn-Jier James Lien, Takeo Kanade, Jeffrey F. Cohn, and C. C. Li. Detection, tracking, and classification of action units in facial expression. *Journal of Robotics and Autonomous System*, 31:131–146, 2000.

[13] Kenji Mase. Recognition of facial expression from optical flow. *IEICE Transactions*, E. 74(10):3474–3483, October 1991.

[14] Y. Moses, D. Reynard, and A. Blake. Determining facial expressions in real time. In *Proc. Of Int. Conf. On Automatic Face and Gesture Recognirion*, pages 332–337, 1995.

[15] M. Rosenblum, Yaser Yacoob, and Larry S. Davis. Human expression recognition from motion using a radial basis function network archtecture. *IEEE Transactions On Neural Network*, 7(5):1121–1138, 1996.

[16] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 20(1):23–38, January 1998.

[17] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database of human faces. In *Proceedings of International Conference on Face and Gesture Recognition*, May, 2002.

[18] M. Singh, G. D. Seyranian, and D.D Hoffman. Parsing silhouettes: the short-cut rule. *Perseption and Psychophysics*, 61(4):636–660, May 1999.

[19] M. Suwa, N. Sugie, and K. Fujimora. A preliminary note on pattern recognition of human emotional expression. In *International Joint Conference on Pattern Recognition*, pages 408–410, 1978.

[20] D. Terzopoulos and K. Waters. Analysis of facial images using physical and anatomical models. In *IEEE International Conference on Computer Vision*, pages 727–732, 1990.

[21] Y.L. Tian, T. Kanade, and J.F. Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 23(2):1–19, February 2001.

[22] Y. Yacoob and L.S. Davis. Recognizing human facial expression from long image sequences using optical flow. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 18(6):636–642, June 1996.

[23] J. Yang, R. Stiefelhagen, U. Meier, and A. Waibel. Real-time face and facial feature tracking and applications. In *Proceedings of Auditory-Visual Speech Processing (AVSP 98), New South Wales, Australia*, 1998.

[24] Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6):893–911, 1999.